Transmission Strategies for Remote Estimation with an Energy Harvesting Sensor

Transmission Strategies for Remote Estimation with an Energy Harvesting Sensor

Ayça Özçelikkale, Tomas McKelvey, Mats Viberg

***Abstract*—We consider the remote estimation of a time-correlated signal using an energy harvesting (EH) sensor. The sensor observes the unknown signal and communicates its observations to a remote fusion center using an amplify-and-forward strategy. We consider the design of optimal power allocation strategies in order to minimize the mean-square error at the fusion center. Contrary to the traditional approaches, the degree of correlation between the signal values constitutes an important aspect of our formulation. We provide the optimal power allocation strategies for a number of illustrative scenarios. We show that the most majorized power allocation strategy, i.e. the power allocation as balanced as possible, is optimal for the cases of circularly wide-sense stationary (c.w.s.s.) signals with a static correlation coefficient, and sampled low-pass c.w.s.s. signals for a static channel. We show that the optimal strategy can be characterized as a water-filling type solution for sampled low-pass c.w.s.s. signals for a fading channel. Motivated by the high-complexity of the numerical solution of the optimization problem, we propose low-complexity policies for the general scenario. Numerical evaluations illustrate the close performance of these low-complexity policies to that of the optimal policies, and demonstrate the effect of the EH constraints and the degree of freedom of the signal.**

I. Introduction

Energy harvesting solutions offer a promising framework for future wireless sensing systems. Instead of completely relying on a fixed battery or power from the grid, nodes with EH capabilities can collect energy from the environment, such as solar power or power from mechanical vibrations. In addition to enabling energy autonomous sensing systems, EH capabilities also offer prolonged network life-times and enhanced mobility for the nodes in the network [1], [2].

One of the key issues in the design of EH systems is the intermittent nature of the energy supply. In a traditional device, the energy that can be used for communications has either a fixed known value for each transmission or there is a total energy constraint. In contrast, for an EH node, the energy available for information transmission depends on the energy used in previous transmissions and the energy that may be available in the future. In such systems, the transmission strategies have to be re-designed in order to ensure reliable and efficient operation in the entire time frame of interest. For instance, at a given instant, an EH node may have to choose between increasing the energy used in the current transmission to increase reliability at that instant or saving the energy for upcoming transmissions due to forecasted poor energy harvesting conditions in the future.

In that respect, the problem of reliable communications with EH nodes have been studied under a broad range of scenarios [1–9]. Capacity of point-to-point Gaussian channels are considered in [3], [4]. Total throughput maximization and transmission time completion problems are investigated in [5], [6]. Multi-user scenarios have been considered, including broadcast channels [7], [8] and multiple-access channels [9]. An overview of these recent advances in EH communication systems is provided in [1], [2]. In contrast to the these works, whose focus is on the reliable communication problem, here we adopt an alternative approach and focus on the estimation aspect of the problem, i.e. recovery of the unknown signal measured by the sensors.

At the moment, the literature on the estimation aspect, in particular investigations on the effect of the possible statistical correlation between the unknown signal values, is quite limited. Previously, the degree of correlation of the unknown signal has been shown to have a substantial impact on the optimum sensor communication strategies without EH constraints [10–13]. In the case of EH systems, only a limited number of works address this issue. Optimal transmission strategies for the estimation of independently identically distributed (i.i.d.) Gaussian sources follow from the findings of [3], [14], [15]. Majorization based arguments of [3] show that energy allocations that are as balanced as possible are optimal for i.i.d. sources. Estimation of i.i.d. sources is considered under a source coding perspective, and an associated 2-D water-filling interpretation is provided in [14]. A water-filling type characterization of optimal solutions for uncoded transmission are provided by [15]. The parameter estimation problems considered in [16], [17] provide insights about the limiting case, where the unknown value is fully spatially correlated across sensors. In particular, a threshold based policy is shown to be optimal under a binary energy allocation strategy [16]. Extensions of this framework, where energy sharing between sensors are possible, is provided in [17]. Investigations in [18–20] provide guidelines for Markov sources. A threshold based strategy is found to be optimal for Markov sources where the sensor transmits if the difference between the current source value and the most recently transmitted value exceeds the threshold [18]. Optimal power allocations for a vector Gaussian Markov source under an unreliable channel with packet erasures is considered in [19]. A characterization of the optimal power allocations for temporally correlated Markov sources is provided in terms of water-filling type solutions under a source-coding framework in [20]. A distributed source

A. Özçelikkale, T. McKelvey and M. Viberg are with Dep. of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden e-mails: {ayca.ozcelikkale, tomas.mckelvey, mats.viberg}@chalmers.se. This work was supported in part by European Union's Horizon 2020 research and innovation programme under grant agreement No 654123.

coding framework for spatially correlated sources is considered in [21], [22].

Here we focus on the estimation of a time-correlated Gaussian signal using an EH sensor. The EH sensor observes the unknown signal and communicates its observations to the remote fusion center under energy harvesting constraints. We consider an amplify-and-forward strategy motivated by the high computational cost of source and channel coding operations; and the fact that EH devices are typically low-power devices that may not have the complex circuitry required for these operations. We note that for estimation of a Gaussian source, uncoded transmission (analog forwarding) is optimal for additive white Gaussian (AWGN) channels under mean-square error without EH constraints [23], [24]. This result has also been extended to the energy harvesting scheme for i.i.d. Gaussian signals in the asymptotic regime [15]. We focus on the problem of optimal power allocation in order to minimize the mean-square error (MSE) over a finite-length horizon at the fusion center. Here we consider a general fading channel scenario whereas an investigation for the static channel case with limited proofs is provided in [25].

We adopt the off-line optimization scheme, where the sensor knows the energy arrivals and the channel gains acausally. Off-line optimization approaches have been investigated for various scenarios, such as point-to-point channels [5], [6], broadcast channels [7], [8] and multiple-access channels [9] under rate based performance criterion as well as for source coding [14], [20], [21] and remote estimation scenarios [16]. From an energy harvesting perspective, these type of approaches are well-suited for scenarios where the energy arrivals can be accurately predicted, such as RF energy harvesting scenarios with dedicated power transfer scheduling as in [26], [27]. Off-line optimization approaches also provide benchmarks to evaluate the fundamental performance limitations for energy harvesting systems and structural guidelines which facilitate possibly sub-optimal but efficient solutions for the general case. Examples for this include the online near-optimal scheme of [28] which uses the off-line directional water-filling solution of [5] and the block transmission scheme of [29] which is motivated by the most-majorized power allocation of [3] optimal for the off-line scheme.

We provide the optimal power allocation strategies for a number of illustrative scenarios. We present water-filling type characterizations of the optimal strategies for uncorrelated sources. These characterizations make use of a time-index dependent threshold, which is a typical property of the EH solutions [5]. For the parameter estimation case, i.e. fully correlated signal scenario, the strategy that only sends the data in the time slots with the most favorable channel conditions is shown to be optimal. We also consider circularly wide-sense stationary signals, which constitute a finite dimensional analog of wide-sense stationary signals [30], [31]. We note that, in general, the components of c.w.s.s. signals are possibly correlated and the calculation of mean-square error requires a matrix inversion as opposed to a direct sum of rate functions as in the case of throughput based formulations [6–8]. Nevertheless, we show that water-filling type characterizations of optimal strategies also hold for sampled low-pass c.w.s.s. signals for fading channels. We also show that the most majorized power allocation strategy, i.e. the power allocation as balanced as possible, is optimal regardless of the degree of correlation in the cases of c.w.s.s. signals with static correlation coefficient and sampled low-pass c.w.s.s. signals for a static channel. Although one may expect that as the signal components become more correlated, strategies that send a low number of signal components with higher power become optimal instead of strategies that allocate power as uniform as possible, the case of static correlation shows that this is not always the case.

These results on c.w.s.s. signals complement the other scenarios where balanced power allocations are found to be optimal, in particular, the i.i.d. sources scenario that follows from the findings of [3] and sensing of two correlated Gaussian variables studied in a rate-distortion framework in [21]. We note that, by definition, the covariance matrices associated with c.w.s.s. signals are circulant [30], [31]. Due to the asymptotic equivalence of sequences of circulant and Toeplitz matrices, (which constitute the covariance matrices of wide-sense stationary signals [31]), our investigations here can be considered as an intermediate step towards understanding limitations imposed by energy harvesting to sensing of wide-sense stationary signals, which is a fundamental signal model in the fields of communications and signal processing.

Motivated by the high complexity of the numerical solution of the optimization problem for the general scenario, we propose a number of low-complexity policies. These policies are based on lower and upper bounds on the mean-square error and provide possibly sub-optimal but nevertheless efficient approaches to the power allocation problem at hand. Numerical evaluations illustrate the close performance of the low-complexity policies to that of the optimal policies, and demonstrate the effect of the energy harvesting constraints and the degree of freedom of the signal on the system performance.

The rest of the paper is organized as follows. We present the problem formulation in Section II. In Section III, the optimal strategies for a number of scenarios are provided. In Section IV, low-complexity strategies for the general case are proposed. In Section V, we present heuristic policies for the fading channel scenario. Numerical evaluations are provided in Section VI. The paper is concluded in Section VII.

Notation: The complex conjugate transpose of a matrix A is denoted by $A^\dagger$. The i^{th} row, k^{th} column element of a matrix A is denoted by $[A]_{ik}$. The positive semi-definite (p.s.d.) partial ordering for Hermitian matrices is denoted by $\succeq$. I_n denotes the identity matrix with $I_n \in \mathbb{C}^{n\times n}$.

II. System Model and Problem Statement

A. Signal Model

The aim of the remote estimation system is to estimate the unknown complex proper zero-mean Gaussian signal $\mathbf{x}$ defined over time as $\mathbf{x} = [x_1, \ldots, x_t, \ldots, x_n] \in \mathbb{C}^{n\times 1}$, $\mathbf{x} \sim \mathcal{CN}(0, K_\mathbf{x})$ with $K_\mathbf{x} = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$, $P_x \triangleq \mathrm{tr}[K_\mathbf{x}]$. We denote the eigenvalue decomposition (EVD) of $K_\mathbf{x}$ as $K_\mathbf{x} = U\Lambda_x U^\dagger$, where $\Lambda_x \in \mathbb{R}^{n\times n}$ is the diagonal matrix of eigenvalues and $U \in \mathbb{C}^{n\times n}$ is a unitary matrix. Let s with $s \leq n$ be

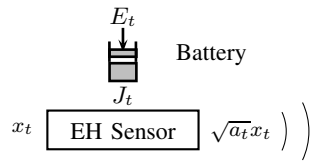


Fig. 1: Energy Harvesting Sensor

the number of non-zero eigenvalues of $K_{\mathbf{x}}$, i.e. rank of $K_{\mathbf{x}}$. Let Ω denote the index set of non-zero eigenvalues. Hence $K_{\mathbf{x}} = U_\Omega \Lambda_{x,s} U_\Omega^\dagger$ is the reduced eigenvalue decomposition of $K_{\mathbf{x}}$ where $\Lambda_{x,s} \in \mathbb{R}^{s\times s}$ is the diagonal matrix of non-zero eigenvalues and $U_\Omega \in \mathbb{C}^{n\times s}$ is the sub-matrix formed by the columns of U corresponding to the non-zero eigenvalues.

B. Sensing and Communications to the Fusion Center

Motivated by the high computational cost of source and channel coding operations, and the fact that typical EH devices are low-power devices which may not have the complex circuity required for these operations, we consider an amplify-and-forward strategy for the sensor similar to [15–17]. As illustrated in Fig. 1, at time slot t, the sensor measures x_t, the unknown signal value at time t and communicates it to the fusion center as follows:

$$y_t = h_t\sqrt{a_t}x_t + w_t, \quad t = 1,\dots,n \tag{1}$$

where $h_t \in \mathbb{C}$, $\sqrt{a_t} \in \mathbb{R}$, $y_t \in \mathbb{C}$ and $w_t \in \mathbb{C}$ denote the channel fading coefficient, the amplification factor adopted by the sensor, the received signal at the fusion center, and the channel noise respectively. Here $\mathbf{w} = [w_1,\dots,w_n] \in \mathbb{C}^{n\times 1}$ is complex proper zero-mean Gaussian with $\mathbf{w} \in \mathbb{C}^{n\times 1} \sim \mathcal{CN}(0, K_{\mathbf{w}})$, $K_{\mathbf{w}} = \sigma_w^2 I_n$, $\sigma_w^2 > 0$.

C. Energy Constraints at the Sensor

The average energy used by the sensor during transmission of x_t can be written as follows [15–17]

$$J_t = \tau\mathbb{E}[||\sqrt{a_k}x_t||^2] = \tau a_t \sigma_{x_t}^2, \tag{2}$$

where the transmit duration is taken as $\tau = 1$ in the rest of the paper. Communications system design under average power constraints have been considered for a wide range of scenarios, including amplify-forward strategy design [10], [11] and linear encoder design [32] without the energy harvesting constraints. Here we consider an amplify-forward scenario under EH constraints. At each time slot t, an energy packet of E_t arrives at the battery. We consider the off-line scheme, where E_t have arbitrary, but known values, during the time frame $t = 1,\dots,n$ [6–9]. The sensor operates under the following energy neutrality conditions

$$\sum_{l=1}^{t} J_l \le \sum_{l=1}^{t} E_l, \qquad t = 1,\dots,n. \tag{3}$$

where the initial energy at the battery is zero. These conditions ensure that the energy used at any time does not exceed the available energy. Here we consider a device with a large enough battery capacity so that no energy packet E_t has to be dropped.

D. Estimation at the Fusion Center

After receiving $\mathbf{y} = [y_1,\dots,y_n] \in \mathbb{C}^{n\times 1}$, the fusion center forms the minimum MSE (MMSE) estimate of $\mathbf{x}$, i.e. $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{y}] = K_{\mathbf{xy}}K_{\mathbf{y}}^{-1}\mathbf{y}$ [33, Ch2]. We have

$$\mathbb{E}[\mathbf{x}\mathbf{y}^\dagger] = K_{\mathbf{xy}} = K_{\mathbf{x}}A^\dagger H^\dagger,$$
$$\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger] = K_{\mathbf{y}} = HAK_{\mathbf{x}}A^\dagger H^\dagger + K_{\mathbf{w}},$$

with $H = \mathrm{diag}(h_t)$, $A = \mathrm{diag}(\sqrt{a_t}) \in \mathbb{R}^{n\times n}$. The resulting MMSE can be expressed as [33, Ch2]

$$\varepsilon(A) = \mathrm{tr}[K_{\mathbf{x}} - K_{\mathbf{xy}}K_{\mathbf{y}}^{-1}K_{\mathbf{xy}}^\dagger]. \tag{4}$$

Hence we have

$$\varepsilon(A) = \mathrm{tr}\left[(\Lambda_{x,s}^{-1} + \gamma U_\Omega^\dagger \,\mathrm{diag}(|h_t|^2 a_t) U_\Omega)^{-1}\right] \tag{5}$$

where $\gamma \triangleq 1/\sigma_w^2$ and (5) follows from (4) and the Sherman-Morrison-Woodbury identity [34]. Here the fusion center uses the source and the noise statistics, including the covariance matrices; and the amplification factors and the channel gains. We note that the same type of later knowledge are needed at the receivers when rate based performance metrics are used [5–9]. We further discuss these points in Section II-E.

We note that by adopting a second-order analysis framework and using the optimum *linear* MMSE filter instead of the MMSE filter at the fusion center, the above error analysis can be also performed under non-Gaussian statistics.

E. Problem Statement

Our goal is to design the optimal transmission strategies in order to minimize the MMSE as follows

$$\min_{A} \quad \varepsilon(A) \tag{6a}$$
$$\text{s.t.} \sum_{l=1}^{t} a_l\sigma_{x_l}^2 \le \sum_{l=1}^{t} E_l, \qquad t = 1,\dots,n-1, \tag{6b}$$
$$\sum_{l=1}^{n} a_l\sigma_{x_l}^2 = E_{tot}, \tag{6c}$$
$$a_t \ge 0, \qquad t = 1,\dots,n, \tag{6d}$$

where the constraints (6b)-(6c) follow from (2), (3) with $E_{tot} \triangleq \sum_{l=1}^{n} E_l$. Since for any optimum strategy all the available energy should be used, (6c) is stated as an equality.

Here we consider a scenario where the sensor knows the energy arrivals and the channel gains for a look-ahead window of size n, i.e. off-line optimization as investigated for a wide-range of scenarios, including rate-based metrics [6–9] and source coding/estimation [14–16], [20], [21]. This type of off-line optimization approaches are suitable for energy harvesting scenarios with dedicated power transfer, for instance as in [26], [27] where wireless power transfer is scheduled a priori. They also provide benchmarks for performance limits of energy harvesting systems and structural guidelines for efficient solutions in the general case. Examples for this include the online near-optimal scheme of [28] utilizing the off-line directional water-filling solution of [5] and the block transmission scheme of [29] motivated by the off-line optimal most-majorized power allocation of [3, Sec.7].

We now discuss the convexity properties of the formulation in (6). The objective function of (6) is a convex function since $\mathrm{tr}[X^{-1}]$ is convex for $X \succ 0$ and $X = \Lambda_{x,s}^{-1} + \gamma U_\Omega^\dagger \mathrm{diag}(|h_t|^2 a_t) U_\Omega$ is an affine function of the optimization variables a_t's. The constraints form convex constraints since they are in the form of linear inequalities and equalities. Hence (6) is a convex formulation and the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality under the assumption of a strictly feasible point. Optimal solutions can be found using the standard numerical optimization tools, such as SDPT3, SeDuMi and CVX [35–37].

III. Optimal Transmission Policies

Here we discuss the structure of the solutions for a number of illustrative scenarios. These results motivate the low-complexity policies proposed in Section IV.

A. Uncorrelated Sources

Here we consider the case where the components of $\mathbf{x}$ are uncorrelated, hence $K_\mathbf{x} = \mathrm{diag}(\sigma_{x_t}^2)$, $\sigma_{x_t}^2 > 0$. The MMSE can then be expressed as follows:

$$\varepsilon(A) = \sum_{t=1}^{n} \frac{\sigma_{x_t}^2}{1+\gamma |h_t|^2 \sigma_{x_t}^2 a_t}. \tag{7}$$

The Lagrangian is given by

$$\mathcal{L} = \sum_{t=1}^{n} \frac{\sigma_{x_t}^2}{1+\gamma |h_t|^2 \sigma_{x_t}^2 a_t} + \sum_{T=1}^{n-1} \eta_T W_T + \nu W_n - \sum_{t=1}^{n} \mu_t a_t, \tag{8}$$

where

$$W_k = \sum_{t=1}^{k} \sigma_{x_t}^2 a_t - \sum_{t=1}^{k} E_t, \qquad 1 \le k \le n \tag{9}$$

Here $\eta_T \in \mathbb{R}$, $\eta_T \ge 0$, $1 \le T \le n-1$, $\nu \in \mathbb{R}$ and $\mu_t \in \mathbb{R}$, $\mu_t \ge 0$, $1 \le t \le n$ are the Lagrange multipliers. Hence together with the feasibility conditions, the KKT conditions can be expressed as follows:

$$-\frac{\gamma |h_t|^2 \sigma_{x_t}^4}{(1+\gamma |h_t|^2 \sigma_{x_t}^2 a_t)^2} + \sum_{T=t}^{n-1} \sigma_{x_t}^2 \eta_T + \sigma_{x_t}^2 \nu + \mu_t = 0, \quad \forall t \tag{10}$$

$$\eta_T W_T = 0, \qquad T = 1, \ldots, n-1 \tag{11}$$

$$\mu_t a_t = 0, \qquad t = 1, \ldots, n \tag{12}$$

Solving the KKT conditions reveals that the optimal a_t can be expressed as

$$a_t = \frac{1}{|h_t|\sqrt{\gamma}} \frac{1}{\sigma_{x_t}^2} \left(\sqrt{\frac{\sigma_{x_t}^2}{\kappa_t}} - \frac{1}{|h_t|\sqrt{\gamma}} \right)^+ \tag{13}$$

where c^+ is defined as $c^+ \triangleq \max(0, c)$ and

$$\kappa_t \triangleq \sum_{T=t}^{n-1} \eta_T + \nu \tag{14}$$

can be interpreted as a time-index dependent threshold, which is a typical property of the EH solutions [5], [14], [15], [20]. We note that optimum values of a_t have the same form and they are tied only through a set of thresholds and the feasibility conditions. These type of solutions are often referred to as "water-filling" solutions. The solutions presented in [5], [14], [15], [20] are some examples from the energy harvesting literature. Other more standard water-filling solutions that do not consider EH constraints include the water-filling solutions for capacity maximization in [38, Ch. 10] and the reverse water-filling solutions for rate-distortion function minimization in [38, Ch. 13].

The solution structure in (13) dictates that x_t is sent over the channel with a non-zero power whenever the a priori uncertainty in this component is relatively large, i.e. $\sigma_{x_t}^2 > \kappa_t/(|h_t|^2\gamma)$. If the a priori uncertainty in this component is relatively small, i.e. this condition is not satisfied, a_t is chosen as $a_t = 0$ and x_t is not sent, hence the energy is saved for future transmissions. We note that here $1/(|h_t|^2\gamma)$ can be interpreted as the effective channel noise-to-signal ratio, hence for a transmission to occur, the a priori signal uncertainty should be above the effective noise-to-signal ratio scaled by κ_t.

We note that optimal strategies become more generous with energy expenditure as time passes for a static channel, i.e. $h_t = 1$. More precisely, we obtain the following:

Lemma 3.1: *Let $H = I_n$. Let t_- and t_+ denote the ordering of two time indices with $1 \le t_- \le t_+ \le n$. Let $\sigma_{x_{t_+}}^2 \ge \sigma_{x_{t_-}}^2 > 0$. Then the following holds: i) $a_{t_+}\sigma_{x_{t_+}}^2 \ge a_{t_-}\sigma_{x_{t_-}}^2$; ii) If $a_{t_-} > 0$, then $a_{t_+} > 0$.*

Proof: We note that $\eta_T \ge 0$, hence we have $\kappa_{t_-} \ge \kappa_{t_+}$ i.e. κ_t is a decreasing function of t. Part (i) follows from $\kappa_{t_-} \ge \kappa_{t_+}$ and $a_t\sigma_{x_t}^2 = \frac{1}{\sqrt{\gamma}}(\sqrt{\frac{\sigma_{x_t}^2}{\kappa_t}} - \frac{1}{\sqrt{\gamma}})^+$. Part (ii) follows from Part (i) with $a_{t_-} > 0$.

Part (i) states that if an energy of $a_t\sigma_{x_t}^2$ has been used before, one will not use less energy for any subsequent component with higher variance. Part (ii) states that if a signal component with a given variance has been sent before (i.e. $a_{t_-} > 0$), all the components with higher variance (i.e. higher uncertainty) will also be sent over the channel in the future.

We now take a closer look to the solution structure in the case where the source is white:

1) White Sources: Here $K_\mathbf{x} = \sigma_x^2 I_n$ by definition. Under $H = I_n$, the MMSE can be expressed as follows:

$$\varepsilon(A) = \sum_{t=1}^{n} \frac{\sigma_x^2}{1+\gamma \sigma_x^2 a_t}. \tag{15}$$

Such sources have been investigated in [15] using the KKT conditions. Here we adopt an alternative approach and illustrate how optimal strategies can be found by adopting the arguments of [3]. More precisely, we note the following:

Definition 3.1: *[39, Ch.1] Let $a = [a_1, \ldots, a_n] \in \mathbb{R}^n$ and $b = [b_1, \ldots, b_n] \in \mathbb{R}^n$. Then a is said to be majorized by b if the following holds:*

$$\sum_{t=1}^{k} a_{[t]} \le \sum_{t=1}^{k} b_{[t]}, \qquad k = 1, \ldots, n-1 \tag{16}$$

$$\sum_{t=1}^{n} a_{[t]} = \sum_{t=1}^{n} b_{[t]} \tag{17}$$

Here $a_{[t]}$ denotes the components of a in decreasing order, i.e. $a_{[1]} \geq, \ldots, \geq a_{[n]}$. This majorization relationship is denoted by $a \prec b$.

Majorization can be interpreted as a measure of how balanced the distribution of the components of vectors are. In particular, the following relationship holds: $\forall a \in \mathbb{R}$: $\bar{a} \prec a \prec \tilde{a}$, where $\bar{a} = (S_a/n)[1, \ldots, 1] \in \mathbb{R}^n$ and $\tilde{a} = [0, \ldots, 0, S_a, 0, \ldots, 0] \in \mathbb{R}^n$ has only one non-zero component, where $S_a = \sum_{t=1}^{n} a_t$. Hence, every vector majorizes the vector that has equal components and has the same total sum, and every vector is majorized by the vector that has only one non-zero component with the same total sum. The following is of interest:

Definition 3.2: *[39, Ch.3] Let $\mathcal{S} \subseteq \mathbb{R}^n$ and $f(\cdot) : \mathcal{S} \to \mathbb{R}$. Then $f(\cdot)$ is said to be Schur-convex on $\mathcal{S}$ if $a \prec b$ on $\mathcal{S}$ implies $f(a) \leq f(b)$.*

Lemma 3.2: *[39, Ch.3] Let $\mathcal{S} \subseteq \mathbb{R}$, and $g(\cdot) : \mathcal{S} \to \mathbb{R}$ be convex. Then $f(a) = \sum_{t=1}^{n} g(a_t)$ is Schur-convex.*

By Lemma 3.2, (15) is Schur-convex since $g(a_t) = \frac{\sigma_x^2}{1+\gamma\sigma_x^2 a_t}$ is a convex function of a_t, $a_t \geq 0$. Hence an optimal solution is given by a_t that is majorized by all feasible power allocations, i.e. the allocation which is as balanced as possible, or alternatively as uniform as possible. Characterization of such solutions have been studied in relation to maximization of the rate function in [3]:

Lemma 3.3: *[3, Thm.3] The power allocation that is majorized by all feasible solutions of* (6b)*,* (6c)*, can be characterized as follows:*

$$\bar{a}_r = \frac{\bar{E}_{\tau_k} - \bar{E}_{\tau_{k-1}}}{\tau_k - \tau_{k-1}}, \quad r = \tau_{k-1} + 1, \ldots, \tau_k \tag{18}$$

$$\tau_k = \arg \min_{r \in \{\tau_{k-1}+1, \ldots, \bar{\tau}\}} \frac{\bar{E}_r - \bar{E}_{\tau_{k-1}}}{r - \tau_{k-1}}, \; k = 2, \ldots, K \tag{19}$$

where $1 \leq r \leq n$, $\tau_1 = 0$ and $\bar{\tau} = \tau_{K+1} = n$, and $1 \leq K \leq n$ is the number of constant power sections.

Here we have adopted the notation $\bar{E}_L = \sum_{t=1}^{L} E_t/\sigma_x^2$, $a_t = \bar{a}_r$ with $r = t, \forall r, t$ for later notational convenience. Hence we obtain the following:

Corollary 3.1: *Let $K_{\mathbf{x}} = \sigma_x^2 I_n$, $H = I_n$. Then* (18)-(19) *provide an optimal solution for* (6)*.*

Proof: The result follows from Schur-convexity of (15).

In the subsequent sections, we will utilize Lemma 3.3 to provide optimal solutions in scenarios even when the source is not white.

B. Parameter Estimation

We now consider the scenario where $K_{\mathbf{x}}$ is of rank 1, hence there is effectively only one random variable to be estimated. We refer to this case as the parameter estimation scenario. In this case, $K_{\mathbf{x}} = U_\Omega \Lambda_{x,1} U_\Omega$ where $U_\Omega \in \mathbb{C}^{n \times 1}$, $\Lambda_{x,1} = P_x$. Let $u_t \in \mathbb{C}$ denote the tth component of U_Ω. The correlation coefficient between x_{t_1} and x_{t_2} is given by

$$\rho_{t_1 t_2} = \frac{\mathbb{E}[x_{t_1} x_{t_2}]}{\sigma_{x_{t_1}} \sigma_{x_{t_2}}} = \frac{P_x u_{t_1} u_{t_2}^\dagger}{(P_x^{1/2}|u_{t_1}|)(P_x^{1/2}|u_{t_2}|)} = \frac{u_{t_1} u_{t_2}^\dagger}{|u_{t_1}||u_{t_2}|}.$$

Hence, $|\rho_{t_1 t_2}| = 1$, $\forall t_1, t_2$. Hence, when $K_{\mathbf{x}}$ is of rank 1, the signal can be said to be fully correlated. The error can be expressed as

$$\varepsilon(A) = \frac{1}{1/P_x + \gamma \sum_{t=1}^{n} |h_t|^2 |u_t|^2 a_t}, \tag{20}$$

$$= \frac{1}{1 + \gamma \sum_{t=1}^{n} |h_t|^2 \sigma_{x_t}^2 a_t} P_x, \tag{21}$$

where we have used $|u_t|^2 P_x = \sigma_{x_t}^2$. Optimal solutions can be characterized as follows:

Lemma 3.4: *An optimum strategy for* (6) *for the parameter estimation case is given by the following recursive procedure:*
i) Initialization: Let $a_t = 0$, $\forall t$. Let $i = 1$; $t^* = 0$.
ii) Let $S_i = [t^* + 1, \ldots, n]$. Let $E_c(t) = \sum_{l=t^*+1}^{t} E_l$, $t \in S_i$.
iii) Let $t^* = \arg\max_{t \in S_i} |h_t|^2$. Then $a_{t^*} = E_c(t^*)/\sigma_{x_{t^*}}^2$.
iv) If $t^* \neq n$, update i as $i = i+1$ and go to Step-ii. Otherwise stop.

The proof is given in Section VIII-A. This procedure sends the data in the most favorable time slots, i.e. the time slots with the highest channel gains, under the energy causality constraints. In particular, in the first iteration, the time slot with the highest gain is determined. Let us refer to this time slot as t_a. Hence in the first iteration, a transmission at t_a with all the energy stored in the battery up to t_a is scheduled (hence no transmission should occur up to t_a). In the next iteration, the time slot with the highest channel gain is found among the time slots after t_a. Let us refer to this time slot as t_b, where $t_b \geq t_a$ by construction. The previous procedure is repeated at t_b; all the energy stored in the battery between time slots t_a and t_b is used for the transmission at t_b and no transmissions should occur in between t_a and t_b. This procedure is repeated until the end of n time steps is reached.

We now focus on the static channel case, i.e. $H = I_n$: Since we have $\sum_{t=1}^{n} \sigma_{x_t}^2 a_t = E_{tot}$ by (6c), evaluating (21) for $H = I_n$ reveals that any feasible strategy is an optimum strategy including the most uniform strategy given by (18)-(19). The optimum error value is given by $(1+\gamma E_{tot})^{-1} P_x$. This result shows that in the case of a fully correlated source and the static channel, the correlation between the signal values can be used to completely compensate for the unreliability of the EH source as long as the total energy that arrives at the sensor after n time steps stays constant.

C. A Lower Bound

We will now consider a lower bound on the performance. In the upcoming sections, we will utilize this lower bound to prove the optimality of some of the proposed policies. We consider the following setting:

$$\varepsilon_{LB} = \min_{A} \quad \varepsilon(A) \tag{22a}$$

$$\text{s.t.} \sum_{l=1}^{n} a_l \sigma_{x_l}^2 = E_{tot} \tag{22b}$$

subject to (6d). Compared to (6), here the energy causality constraints are ignored and only the total energy constraint is imposed. Hence (22) forms a relaxation of (6) and the optimum value of (22) provides a lower bound for the optimum value of (6). We also note that this scenario can be interpreted as fixed battery scenario where a total energy of E_{tot} is available for usage over n time slots. Such scenarios have been studied in distributed estimation scenarios under different assumptions [10], [11].

Let $H = I_n$. To find an analytical expression, we focus on the case where $\Lambda_{x,s}$ is of the form $\Lambda_{x,s} = \frac{P_x}{s} I_s$, i.e. the non-zero eigenvalues are all equal. This type of models have been used to represent signal families with a low degree of freedom in various signal applications, for instance as a sparse signal model in the compressive sensing literature [13], [40]. We obtain the following result for ε_{LB}:

Lemma 3.5: *Let $\Lambda_{x,s} = (P_x/s) I_s$, $H = I_n$. Then $a_t = E_{tot}/P_x$, $\forall t$ is an optimum strategy for* (22)*. The optimal value is given by $\varepsilon_{LB} = \frac{1}{1+\gamma E_{tot}/s} P_x$.*

The proof is presented in Section VIII-B. Hence, whenever $a_t = E_{tot}/P_x$ is a feasible allocation for (6), it is also an optimal strategy. More precisely, we obtain the following result:

Corollary 3.2: *Let $\Lambda_{x,s} = (P_x/s) I_s$, $H = I_n$. If $\frac{1}{P_x}\sum_{l=1}^{t} \sigma_{x_l}^2 \leq \frac{1}{E_{tot}} \sum_{l=1}^{t} E_l$, $\forall t$, then $a_t = E_{tot}/P_x$ is an optimum strategy for* (6) *with the optimal value $\frac{1}{1+\gamma E_{tot}/s} P_x$.*

A constant energy arrival scenario where the conditions of Corollary 3.2 are satisfied is discussed in Section III-D.

D. Circularly Wide-Sense Stationary Signals

We now focus on the c.w.s.s signals, which constitute a finite dimensional analog of wide-sense stationary signals [30], [31]. By definition, the covariance matrix associated with c.w.s.s. signals is circulant, i.e. the matrix is determined by its first row as $[K_{\mathbf{x}}]_{tk} = [K_1]_{\text{mod}_n(k-t)}$, where $K_1 \in \mathbb{C}^{1\times n}$ is the first row of $K_{\mathbf{x}}$ [30], [31].

Due to the asymptotic equivalence of sequences of circulant and Toeplitz matrices, (which constitute the covariance matrices of wide-sense stationary signals [31]), our investigations here can be considered as an intermediate step towards understanding limitations imposed by energy harvesting to sensing of wide-sense stationary signals, which is a fundamental signal model in the fields of communications and signal processing. In particular, one method for computation of the estimation error of a wide-sense stationary discrete time signal is to consider finite sections of the signal with increasing length. Covariance matrices of these finite length signals are given by a sequence of Toeplitz matrices. Such a computation of the estimation error requires evaluations of matrix operations on Toeplitz matrices. Due to the fact that the unitary transform in the eigenvalue decomposition of circulant matrices is always given by the DFT matrix, matrix operations are relatively simple when dealing with circulant matrices compared to Toeplitz matrices [31]. In contrast, there is no fixed unitary transform associated with finite sections of Toeplitz matrices. Nevertheless, using the asymptotic equivalence of Toeplitz and circulant matrices, one may evaluate the estimation error for w.s.s. signals [31], as illustrated in [31] without EH constraints.

Due to stationarity, we have $\sigma_{x_t}^2 = \sigma_x^2 = P_x/n, \forall t$. The unitary matrix U in the EVD of $K_{\mathbf{x}}$ for a circularly wide-sense stationary signal is given by the DFT matrix [30], [31]. Let F^n denote the DFT matrix of size $n \times n$, i.e. $[F^n]_{tk} = (1/\sqrt{n}) \exp(-j\frac{2\pi}{n}(t-1)(k-1))$, $1 \leq t,k \leq n$, where $j = \sqrt{-1}$. Hence, the reduced EVD of $K_{\mathbf{x}}$ is given by $K_{\mathbf{x}} = F_\Omega^n \Lambda_{x,s} {F_\Omega^n}^\dagger$, where $\Lambda_{x,s} = \text{diag}(\lambda_k) \in \mathbb{R}^{s\times s}$ and $F_\Omega^n \in \mathbb{C}^{n\times s}$ is the matrix that consists of s columns of F^n corresponding to non-zero eigenvalues.

Constant energy arrival scheme with $\Lambda_{x,s} = (P_x/s) I_s$: To gain some insight into the optimal power allocations in the case of c.w.s.s. signals, we now consider the case with $\Lambda_{x,s} = (P_x/s) I_s$ under constant energy arrival scheme, i.e. $E_t = E$, $\forall t$. We observe the following: Due to $\sigma_{x_t}^2 = P_x/n, \forall t$, the conditions of Corollary 3.2 are satisfied for this scenario. Hence the lower bound presented in Lemma 3.5 is achieved even under the energy causality constraints in such scenarios.

We now go back to general c.w.s.s. scenario with arbitrary E_t's. We obtain the following, which we will utilize later:

Lemma 3.6: *Let $H = I_n$. Let $e_i \in \mathbb{R}^n$, $1 \leq i \leq n$ denote the i^{th} unit vector. Let the EVD of $K_{\mathbf{x}}$ be given by $K_{\mathbf{x}} = F^n \Lambda_x {F^n}^\dagger$ with $\Lambda_x = \beta I_n + \alpha e_i e_i^\dagger$ with $-\beta < \alpha$, $\beta > 0$, $\alpha, \beta \in \mathbb{R}$. Then* (18)-(19) *is an optimal strategy for* (6)*.*

The proof is given in Section VIII-C. This eigenvalue distribution model covers a number of signal families with appealing interpretations. We now identify two such cases, i.e. almost white sources and sources with static correlation coefficient.

1) Almost White Sources: When x_t is white, we have $K_{\mathbf{x}} = \sigma_x^2 I_n$. Hence the EVD of $K_{\mathbf{x}}$ is given by $K_{\mathbf{x}} = U \Lambda_x U^\dagger$ with $\Lambda_x = \sigma_x^2 I_n$, where U is an arbitrary unitary matrix since $UU^\dagger = I_n$ for all unitary matrices. Motivated by this, we refer to the case where $\Lambda_x \propto I_n - \epsilon e_j e_j^\dagger$, $0 < \epsilon < 1$ as an *almost white* source.

We obtain the following result as a direct corollary to Lemma 3.6: Let $H = I_n$. Let $\mathbf{x}$ be almost white with $K_{\mathbf{x}} = F^n \Lambda_x {F^n}^\dagger$, $\Lambda_x = I_n - \epsilon e_j e_j^\dagger$, $0 < \epsilon < 1$. Then (18)-(19) is an optimal strategy for (6). This result shows that even when the source is not exactly white but only close to being white as defined above, the most uniform feasible allocation is still an optimal solution.

2) Static Correlation Coefficient: We now consider the family of signals whose covariance matrix has the following form

$$K(\rho) = \frac{P_x}{n} \begin{bmatrix} 1 & \rho & \dots & \rho \\ \dots & \dots & \dots & \\ \rho & \dots & \dots & 1 \end{bmatrix}, \tag{23}$$

where $K(\rho) \in \mathbb{R}^{n\times n}$, $n \geq 2$, $0 \leq |\rho| \leq 1$, $\rho \in \mathbb{R}$. Hence, the correlation coefficient between x_i and x_j, $i \neq j$ does not depend on i, j. We note that for $K(\rho)$ to be a valid covariance

matrix, it should be positive semi-definite, i.e. $K(\rho) \succeq 0$. Hence ρ should also satisfy $\rho(n-1)+1 \geq 0$. This result is proven alongside with the result for optimal strategies in Lemma 3.7.

We obtain the following result for optimal strategies:

Lemma 3.7: *Let $H = I_n$ and $K_{\mathbf{x}} = K(\rho)$, $\rho \in \mathbb{R}$, $0 \leq |\rho| \leq 1$, $\rho(n-1)+1 \geq 0$. Then* (18)-(19) *is an optimal strategy for* (6)*.*

Proof: Let v be the first row of $K_{\mathbf{x}}$, i.e. $v = (P_x/n)[1, \rho, \ldots, \rho] \in \mathbb{C}^n$. Let $z = [z_1, \ldots, z_n] = [\lambda_1, \ldots, \lambda_n] \in \mathbb{R}^n$ be the vector of eigenvalues. The relationship between the eigenvalues and the first row of a circulant matrix is given by $z = \sqrt{n} F^n v$ [31, Ch.3]. Hence we obtain $z_1 = (P_x/n)(\rho(n-1)+1)$ and $z_i = (P_x/n)(1-\rho)$, $2 \leq i \leq n$. Thus, Lemma 3.6 applies and (18)-(19) is an optimal strategy. We note that to have $z_i \geq 0$, $\forall i$, we should have $\rho \leq 1$ and $\rho(n-1)+1 \geq 0$. Since a Hermitian matrix is positive semi-definite if and only if all of its eigenvalues are non-negative, the conditions $\rho \leq 1$ and $\rho(n-1)+1 \geq 0$ are necessary and sufficient for $K_{\mathbf{x}}$ to be positive semi-definite and a valid covariance matrix. □

Remark 3.1: *Regardless of the value of ρ, i.e. the level of statistical dependency of the signal components, the strategy that allocates the power as balanced as possible is an optimal strategy.*

Although one may expect that as the signal components become more correlated, strategies that send a low number of signal components with higher power become optimal instead of strategies that allocate power as uniform as possible, Lemma 3.7 shows that this is not always the case and uniform power allocation strategies may continue to be optimal. These results complement the other scenarios where such allocations are found to be optimal, in particular the i.i.d. sources scenario that follows from the findings of [3] as discussed in Section III-A1 and the sensing of two correlated Gaussian variables studied in a distributed source coding framework in [21, Prop.3].

3) Low-Pass Signals: Let $n/s \in \mathbb{Z}$. Let us order the eigenvalues of $K_{\mathbf{x}}$ so that λ_k denotes the eigenvalue that corresponds to the eigenvector in the kth column of F^n, where F^n is as defined above. Here we consider low-pass signals, i.e. signals for which $\Omega = \{1, \ldots, s\}$, and $\lambda_1 =, \ldots, = \lambda_s = P_x/s$, and the rest are zero. Hence we have $K_{\mathbf{x}} = F^n_\Omega \Lambda_x F^{n\dagger}_\Omega$, $\Lambda_x = (P_x/s) I_s$.

Similar to their deterministic counterparts, given $\sigma_w^2 = 0$, low-pass c.w.s.s. signals can be recovered from their equidistant samples with zero mean-square error when the number of samples is larger than s, or equivalently the spacing between the samples satisfies $\Delta \leq n/s$ [13]. Motivated by this, we consider communication strategies that send one out of every $\Delta = n/s$ samples, i.e. strategies in the form of

$$a_t = \begin{cases} \geq 0 & \text{if } t = \Delta r + t_d + 1, \quad 0 \leq r \leq m-1 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where $m = n/\Delta$ is the number of samples sent, and $t_d \in 0, \ldots, \Delta - 1$, the initial delay before sending the first data, is fixed.

We now consider the error associated with the scenario where the sensor only sends these equidistant samples to the fusion center. Let $f_n = \exp(-j\frac{2\pi}{n})$. Here, F^n_Ω consists of the first s columns of F^n. Hence, equidistantly row sampled F^n_Ω can be associated with the DFT matrix of size $s \times s$, F^s, as follows

$$[F^n_\Omega]_{(n/s)r+t_d+1,k+1} = \frac{1}{\sqrt{n}} f_n^{((n/s)r+t_d)k}, \tag{25}$$

$$= \frac{1}{\sqrt{n}} f_s^{rk} f_n^{t_d k}, \tag{26}$$

$$= \sqrt{\frac{s}{n}} [F^s]_{r+1,k+1} f_n^{t_d k}, \tag{27}$$

where $0 \leq k \leq s-1$, $0 \leq r \leq s-1$. Let $D = \mathrm{diag}(d_k) \in \mathbb{C}^{s \times s}$, $d_k = f_n^{t_d k}$. Let $\bar{a}_r \triangleq a_{\Delta r + t_d + 1}$. The error can be expressed as follows

$$\varepsilon(\bar{A}) = \mathrm{tr}[(\frac{s}{P_x} I_s + \gamma \frac{s}{n} D F^{s\dagger} \bar{H}^\dagger \bar{A}^\dagger \bar{A} \bar{H} F^s D)^{-1}], \tag{28}$$

$$= \mathrm{tr}[(\frac{s}{P_x} I_s + \gamma \frac{s}{n} \bar{H}^\dagger \bar{A}^\dagger \bar{A} \bar{H})^{-1}], \tag{29}$$

$$= \sum_{r=0}^{s-1} \frac{1}{\frac{s}{P_x} + \frac{s}{n} \gamma \bar{a}_r |\bar{h}_r|^2}, \tag{30}$$

$$= \sum_{r=0}^{s-1} \frac{1}{1 + \gamma \bar{a}_r \sigma_x^2 |\bar{h}_r|^2} \frac{P_x}{s}, \tag{31}$$

where $\bar{A} = \mathrm{diag}(\sqrt{\bar{a}_r}) \in \mathbb{R}^{s \times s}$ and $\bar{H} = \mathrm{diag}(\bar{h}_r) \in \mathbb{R}^{s \times s}$, $\bar{h}_r = h_{\Delta r + t_d + 1}$. Here, (29) follows from the fact that F^s and D are unitary matrices. In (31), we have used the fact that $\sigma_x^2 = P_x/n$. Hence under the equidistant sampling strategy of (24), (6) can be equivalently expressed as

$$\min_{\bar{a}_r} \quad \sum_{r=0}^{s-1} \frac{1}{1 + \gamma \bar{a}_r \sigma_x^2 |\bar{h}_r|^2} \tag{32}$$

subject to $\sum_{r=0}^{t} \bar{a}_r \sigma_x^2 \leq \sum_{r=0}^{t} \bar{\bar{E}}_r$, $\quad t = 0, \ldots, s-2$ and $\sum_{r=0}^{s-1} \bar{a}_r \sigma_x^2 = E_{tot}$ and $\bar{a}_r \geq 0$. Here $\bar{\bar{E}}_r = \sum_{t=t_0}^{\Delta r + t_d + 1} E_t$ with $t_0 = \max(0, \Delta(r-1) + t_d + 2)$.

Remark 3.2: *We observe that* (32) *and the objective function of Section III-A, i.e. the error expression in* (7)*, have the same form. Hence with appropriate notational modifications, the water-filling type characterization of optimal power allocations provided by* (13) *also applies to* (32)*.*

We now consider the static channel case, i.e. $H = I_n$. We obtain the following result:

Lemma 3.8: *Let $H = I_n$, $\Delta = n/s$, $0 \leq t_d \leq \Delta - 1$. An optimal strategy for* (6) *under the setting in* (24)*, i.e. an optimal strategy for* (32)*, is given by* (18)-(19) *with $\bar{a}_r \triangleq a_{\Delta r + t_d + 1}$, $\bar{E}_r = \sum_{t=1}^{\Delta r + t_d + 1} E_t / \sigma_x^2$ and $\tau_1 = 0$, $\bar{\tau} = \tau_{K+1} = s$, and $1 \leq K \leq s$.*

Proof: By (31), under $h_t = 1$, the error can be expressed as $\varepsilon(\bar{A}) = \sum_{r=0}^{s-1} \frac{1}{1 + \gamma \bar{a}_r \sigma_x^2} \frac{P_x}{s}$. Due to Lemma 3.2, this is a Schur-convex function. The result then follows from Lemma 3.3. □

This strategy allocates the power as uniformly as possible among the s samples sent. Hence the most balanced feasible power allocation is an optimum strategy for a sampled low-pass c.w.s.s. signal.

The equidistant sampling strategy can also provide optimal solutions for the general scenario of (6) even when the equidistant sampling constraint is not imposed to achievable sensor strategies:

Corollary 3.3: *Let $H = I_n$, $\Delta = n/s$, $0 \leq t_d \leq \Delta - 1$. If $\bar{a}_r = E_{tot}/(s\sigma_x^2)$, $\bar{a}_r \triangleq a_{\Delta r + t_d + 1}$ is feasible for* (6)*, it is an optimal strategy for* (6) *with an optimum error value of* $\varepsilon_{LB} = \frac{1}{1+\gamma E_{tot}/s} P_x$.

Proof: By (31) and $\bar{a}_r = E_{tot}/(s\sigma_x^2)$, the error can be expressed as $= \frac{1}{1+\gamma \frac{E_{tot}}{s}} P_x$. We observe that the lower bound in Lemma 3.5 is achieved, hence $\bar{a}_r$ is an optimal strategy. □

Hence, if exactly uniform power allocation over equidistant samples is feasible, sending equidistant samples is an optimal solution for c.w.s.s. signals for the general scenario in (6) under static channel.

In general, there may be more than one optimal strategy for (6). We now provide an example for low-pass c.w.s.s. signals. Let us consider $E_t = E \; \forall t$ for a static channel. In this scenario, both of the following power allocations are optimal: i) S_{ua}: uniform power allocation over all the components, i.e. $a_t = E_{tot}/(n\sigma_x^2) = E/\sigma_x^2$, $\forall t$; ii) S_{ue}: uniform allocation over the equidistant samples, i.e. $\bar{a}_r = nE/(s\sigma_x^2)$, $\Delta = n/s$, $t_d = \Delta - 1$. Here optimality of S_{ua} and S_{ue} follow from Corollary 3.2 and Corollary 3.3, respectively.

E. Discussions

Most majorized solutions play a central role in the above investigations. Here complexity concerns constitute an important motivating factor. Another motivating point is the fact that approaches that try to mimic the most majorized solution, i.e. approaches that allocate power as uniformly as possible, are used quite commonly as practical heuristic approaches. Hence, determining in which scenarios this approach is optimal is of interest, as done above.

In the previous sections, we have presented various scenarios where the objective function is Schur-convex and the most majorized solution is an optimal solution. Nevertheless, we note that Schur-convexity is not a necessary condition for the optimal solution to be the most majorized one. To illustrate this point, we note the following example where Schur-convexity is not satisfied but the optimal allocation is the most majorized one:

Let $n = 2$, $|h_1| = |h_2| = 1$, $\gamma = 1$, $\sigma_{x_1}^2 > \sigma_{x_2}^2$, and $E_1 \leq E_2$. Let x_1 and x_2 be uncorrelated. By (7), the objective function can be expressed as

$$\varepsilon = \frac{\sigma_{x_1}^2}{1 + J_1} + \frac{\sigma_{x_2}^2}{1 + J_2}, \tag{33}$$

where $J_i = \sigma_{x_i}^2 a_i$, $i = 1, 2$. The energy harvesting constraints can be expressed as $J_1 \leq E_1$ and $J_1 + J_2 = E_1 + E_2$. Since ε is not a permutation symmetric function of J_i's, and permutation symmetry is a necessary condition for Schur-convexity [39, Thm. A4], ε is not a Schur-convex function of power allocations.

Evaluating (33) reveals that one should allocate as much power as possible to J_1. In particular, one may parametrize the power allocations as $J_1 = E_m - E_d$, $J_2 = E_m + E_d$, where $E_m \triangleq (J_1 + J_2)/2 = (E_1 + E_2)/2$ and $E_d \triangleq (J_2 - J_1)/2$ with $E_d \geq (E_2 - E_1)/2 \geq 0$ due to $E_1 \leq E_2$, $J_1 \leq E_1$ and $J_1 + J_2 = E_1 + E_2$. The objective function in (33) can be written as $\frac{\sigma_{x_1}^2 + \sigma_{x_2}^2 + E_m(\sigma_{x_1}^2 + \sigma_{x_2}^2) + (\sigma_{x_1}^2 - \sigma_{x_2}^2)E_d}{1 + 2E_m + E_m^2 - E_d^2}$. Since $\sigma_{x_1}^2 > \sigma_{x_2}^2$, this function is minimized by the power allocation with the smallest feasible E_d value, which is given by $(E_2 - E_1)/2$. Hence the optimal solution is in the form $J_1 = E_1$, $J_2 = E_2$. This is exactly the most majorized solution under these energy arrivals. Hence optimal power allocation is the most majorized one even if the objective function is not Schur-convex.

IV. Low-Complexity Transmission Policies

We now propose a number of heuristic schemes. These schemes provide possibly sub-optimal but nevertheless low-complexity schemes. We illustrate the performance of these schemes in Section VI.

The objective function in the optimization formulation in (6) includes a matrix inverse which leads to a computationally challenging optimization formulation. Standard numerical optimization tools, such as SDPT3, SeDuMi and CVX [35–37] convert the problem into a semi-definite programming problem, whose computational complexity is in the order of $O(n^{4.5})$ using an interior-point method [41]. Due to this high computational complexity, it is of interest to find schemes which avoid the matrix inverse in (5). In particular, we consider the following upper bound

$$\varepsilon(A) \leq \sum_{t=1}^{n} \frac{\sigma_{x_t}^2}{1 + \gamma |h_t|^2 \sigma_{x_t}^2 a_t}, \tag{34}$$

where the inequality follows from the fact that the right hand side of (34) is the error of the scheme where the possible correlation between the signal values are ignored. In particular, we observe that $\varepsilon(A) = \sum_{t=1}^{n} \mathbb{E}[|x_t - \mathbb{E}[x_t|\mathbf{y}]|^2] \leq \sum_{t=1}^{n} \mathbb{E}[|x_t - \mathbb{E}[x_t|y_t]|^2] = \sum_{t=1}^{n} \frac{\sigma_{x_t}^2}{1 + \gamma |h_t|^2 \sigma_{x_t}^2 a_t}$. Here the first equation is the standard MMSE expression [33, Ch2]. The inequality follows from the fact that $\mathbb{E}[|x_t - \mathbb{E}[x_t|\mathbf{y}]|^2] \leq \mathbb{E}[|x_t - \mathbb{E}[x_t|y_t]|^2]$, where the right-hand side is the mean-square error associated with estimating x_t using only y_t and the left-hand side is the error of estimating x_t using the larger set $\mathbf{y} = [y_1, \ldots, y_n]$.

Utilizing the fact that the bound in (34) couples the optimization variables only through a summation, we propose block based minimization of this upper bound. Let $1 \leq l_w \leq n \in \mathbb{Z}$ with $n/l_w \in \mathbb{Z}$ be the block size. Let $t_i = (i-1)l_w + 1$. At time index t_i, $i = 1, \ldots, n/l_w$, the sensor looks ahead l_w time steps and designs the following strategy:

$$\min_{a_{t_i}, \ldots, a_{t_{i+1}-1}} \quad \sum_{t=t_i}^{t_{i+1}-1} \frac{\sigma_{x_t}^2}{1 + \gamma |h_t|^2 \sigma_{x_t}^2 a_t} \tag{35a}$$

$$\text{s.t.} \quad \sum_{l=t_i}^{t} a_l \sigma_{x_l}^2 \leq \sum_{l=t_i}^{t} E_l, \qquad t = t_i, \ldots, t_{i+1} - 2, \tag{35b}$$

$$\sum_{l=t_i}^{t_{i+1}-1} a_l \sigma_{x_l}^2 = \sum_{l=t_i}^{t_{i+1}-1} E_l, \tag{35c}$$

The overall strategy a_t, $\forall t$ is obtained by solving (35) over n/l_w non-overlapping windows. We observe that any solution found by this approach is a feasible solution for (6). We note that here the main gain in computational complexity is due to using the upper bound. Nevertheless, schemes with $l_w < n$ are of interest, since these need less knowledge on system conditions, for instance future energy arrivals. The performance of (35) together with a discussion of numerical efficiency is presented in Section VI.

We now focus on the case where the non-zero eigenvalues are equal, i.e. $\Lambda_x = (P_x/s)I_s$. We consider the following lower bound:

Lemma 4.1: *Let* $\Lambda_x = \frac{P_x}{s} I_s$. *The following holds:*

$$\varepsilon(A) \geq \frac{P_x}{s}\left(\sum_{t=1}^{n} \frac{1}{1+\gamma|h_t|^2 a_t \sigma_{x_t}^2} + s - n\right). \tag{36}$$

The proof is given in Section VIII-D. We observe that this bound also avoids the matrix inverse in the optimization formulation. Hence, we propose block based minimization of right-hand side of (36) as a heuristic strategy as follows

$$\min_{a_{t_i},\ldots,a_{t_{i+1}-1}} \quad \sum_{t=t_i}^{t_{i+1}-1} \frac{1}{1+\gamma|h_t|^2 a_t \sigma_{x_t}^2} \tag{37}$$

subject to (35b), (35c). We observe that for the static channel case, by Lemma 3.2, the objective function is Schur-convex and the optimal strategies are given by the allocation that makes $a_t\sigma_{x_t}^2$ distribution as balanced as possible. Hence the solutions follow the characterization provided by (18)-(19) with appropriate notational modifications. In particular for $l_w = n$, we will have $\bar{a}_r = a_r\sigma_{x_r}^2$ and $\bar{E}_r = \sum_{t=1}^{r} \bar{E}_t$. We note that in the general fading channel case, there is no known explicit solution and (37) should be solved numerically. This can be done, for instance, by using the off-the-shelf numerical optimization solvers or using a tailored numerical solution for the KKT conditions of Section III-A.

We observe that for c.w.s.s. signals (and other signal models with $\sigma_{x_t}^2 = \sigma_x^2 = P_x/n$), the upper bound given by (34) and the lower bound provided by (36) have the same form, apart from some scaling factors and additive terms that do not depend on a_t. Hence, the error performance is bounded as follows:

$$(\varepsilon_B + s - n)\frac{P_x}{s} \leq \varepsilon(A) \leq \varepsilon_B \frac{P_x}{n}, \tag{38}$$

where ε_B is defined as $\varepsilon_B \triangleq \sum_{t=1}^{n} \frac{1}{1+\gamma|h_t|^2 a_t \frac{P_x}{n}}$. For a given ε_B, the gap between the upper and lower bounds becomes smaller as the gap between s and n decreases. This is consistent with the fact that as s gets closer to n, the signal can be said to be more close to an uncorrelated source. In the limiting case of $s = n$, the bounds are equal as expected, since the inequalities that give rise to both the upper and lower bounds hold with equality in the uncorrelated case.

V. Heuristic Policies under Online Knowledge of Channel Fade Levels

We now focus on the effect of unknown channel coefficients on the error performance. In practice, estimation of channel coefficients are done through pilot signals, hence long term channel coefficient estimation is not practical. We assume that channel coefficients and energy arrivals are i.i.d. over time and consider the following approaches for varying levels of channel state information at the sensor:

Adaptive Policy: We assume that at time t, the channel coefficients up to time t, i.e. $h_1, \ldots, h_t$, are known whereas only statistical knowledge for the future coefficients $h_{t+1}, \ldots, h_n$ are available. Let $B_k = \sum_{l=1}^{k} E_l - \sum_{l=1}^{k} a_l\sigma_{x_l}^2$ denote the energy at the battery at the end of time slot k. At time step t, a_t is found by setting $a_t = a_t^{(t)}$, where $a_t^{(t)}$ is found by solving the following optimization problem:

$$\min_{a_t^{(t)},\ldots,a_n^{(t)}} \quad \frac{\sigma_{x_t}^2}{1+\gamma|h_t|^2\sigma_{x_t}^2 a_t^{(t)}} + \sum_{l=t+1}^{n} \frac{\sigma_{x_l}^2}{1+\gamma\mathbb{E}[|h_l|^2]\sigma_{x_l}^2 a_l^{(t)}} \tag{39}$$

subject to $a_k^{(t)} \geq 0$ and $\sum_{l=t}^{k} a_l^{(t)}\sigma_{x_l}^2 \leq \sum_{l=t}^{k} E_l + B_{t-1}$, $\forall k$ such that $n \geq k \geq t$. At each time step, a_t is found by using the current fading coefficient and the mean of the fade level values for the future. This procedure is repeated at each time step. This policy utilizes (34) and it is partially motivated by the promising numerical performance of the policies based on (34) for the known channel coefficients case, which is illustrated in Section VI. We note that due to usage of mean value for the future channel coefficients, (39) is no longer an upper bound for the mean-square error.

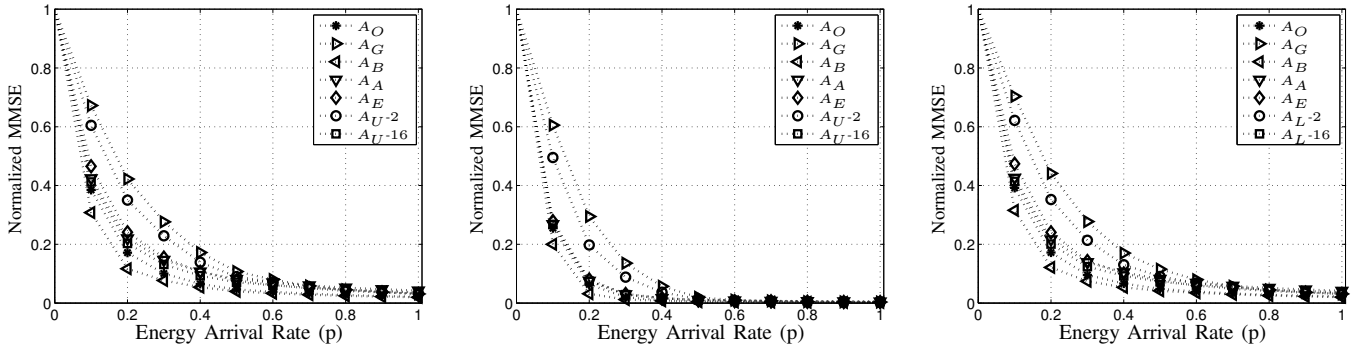
Balancing Policy: Here a design strategy that is completely independent of the channel state information is considered. At each time step $t < n$, a_t is set as

$$a_t = \frac{1}{\sigma_{x_t}^2}\min\left(B_{t-1} + E_t, \mathbb{E}[E_t]\right),$$

where $B_{t-1} + E_t$ is the energy available for usage at time slot t. At the last step, we have $a_n = \frac{1}{\sigma_{x_n}^2}(B_{n-1} + E_n)$. Hence whenever possible, the amount spent is set to the average energy rate, except the last time step where all the available energy is used. This policy is a heuristic approach for balancing the energy allocated to each component.

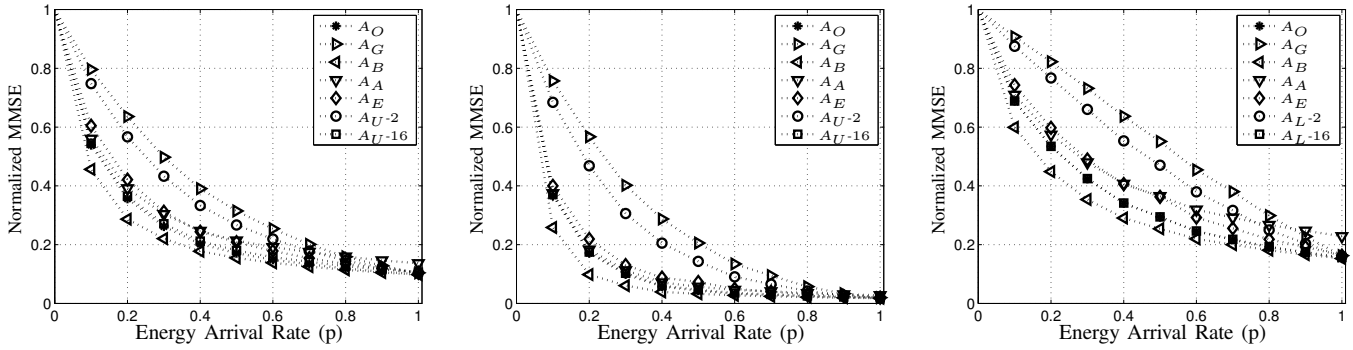
VI. Numerical Results

We now present the numerical evaluations. Let $n=16$, $s=4, 14$, $P_x = n$, $\sigma_w^2 = 0.1$, $\Lambda_{x,s} = \frac{P_x}{\mathrm{tr}[\Lambda]}\Lambda$, $\Lambda = \mathrm{diag}(\alpha_k)$, $\alpha_k = 0.7^k, 0 \leq k \leq s-1$. The unitary matrix U is drawn from the uniform (Haar) unitary matrix distribution [42] and fixed throughout the experiments unless otherwise stated. We denote this unitary matrix with U_a and the DFT matrix with U_f. The energy arrivals are generated with $E_t = \delta_t E_0$, $E_0 = 1$, where δ_t's are i.i.d. Bernoulli with probability of success p, $0 \leq p \leq 1$. We generate h_t as i.i.d. complex proper Gaussian with $h_t \in \mathbb{C}$, $h_t \sim \mathcal{CN}(0, \sigma_h^2)$, $\sigma_h^2 \in \{1, 10\}$. The average error over $N_{sim} = 500$ realizations are reported. The error is normalized as ε/P_x. We refer to ${\sigma_h}^2/{\sigma_w}^2$ as the channel signal-to-noise ratio (SNR). The solutions provided by (6), (35) and (37) are denoted by A_O, A_U-l_w, A_L-l_w, respectively. The greedy approach where the energy is spent as soon as it arrives is denoted by A_G and the lower bound in (22) that ignores the energy neutrality conditions is denoted by A_B. The



(a) $U = U_a$, $\sigma_h^2 = 1$ (b) $U = U_a$, $\sigma_h^2 = 10$ (c) c.w.s.s. ($U = U_f$), $\sigma_h^2 = 1$

Fig. 2: Normalized MMSE versus energy arrival rate, $s = 4$



(a) $U = U_a$, $\sigma_h^2 = 1$ (b) $U = U_a$, $\sigma_h^2 = 10$ (c) c.w.s.s. ($U = U_f$), $\sigma_h^2 = 1$

Fig. 3: Normalized MMSE versus energy arrival rate, $s = 14$

adaptive and the balancing policy of Section V are denoted by A_A and A_E, respectively.

The error versus energy arrival rate curves are presented in Fig. 2 and Fig. 3, for $s = 4$ and $s = 16$, respectively. As expected, due to the low degree of freedom of the signal and the possible high correlation between the signal values, we observe that it is possible to obtain lower error values in Fig. 2 compared to Fig. 3 when the corresponding scenarios in the sub-figures are compared. In both scenarios, the low-complexity scheme with $l_w = n$, A_U-n, is remarkably successful. In particular, the performance of A_O and A_U-n are almost indistinguishable from each other in Fig. 3 whereas there exists a performance gap in the case of low channel SNR and for the signal with low degree of freedom in Fig. 2a and Fig. 2c. This is supported by the need to leverage possible correlation structure in the signal under possibly unfavorable channel conditions. In the case of Fig. 3 the close-to-optimal performance of A_U-n is supported by the relative closeness of the source to an uncorrelated source due to the relatively high degree of freedom provided by $s = 14$. We note that despite this close average performance, the performance gap may be relatively significant for some realizations, and the power allocations provided by A_O and A_U-n may be different. We illustrate these points later in this section.

The error versus energy arrival rate curves for the c.w.s.s. scenarios are presented in Fig. 2c and Fig. 3c, for $s = 4$ and $s = 14$, respectively. Here we have considered the flat eigenvalue distribution scenario with $\Lambda_{x,s} = \frac{P_x}{\mathrm{tr}[\Lambda]}\Lambda$, $\Lambda = \mathrm{diag}(\alpha_k)$, $\alpha_k = 1, 0 \leq k \leq s-1$ so that A_L-l_w applies. The performance of the low-complexity policies A_U-l_w and A_L-l_w are very close, hence we only present the performance of A_L-l_w to avoid clutter in the figures. We observe that again with small s, it is possible to obtain lower error values. Similar to A_U-n, the performance of A_L-n is close to the performance of optimal policies.

We now discuss the performance of the policies that do not require the knowledge of future channel coefficients, i.e. A_G, A_A and A_E. Compared to the performance of the greedy policy A_G, performances of the proposed heuristic policies A_A and A_E are observed to be quite close to the performance of the offline policy A_O. We further discuss the performance of A_A in terms of its gap with A_O below.

We now take a closer look at the performance gap between the optimal policies and the low-complexity policies. Let e_X denote the error associated with the strategy A_X. Let us denote the error gap as $e_G = e_X - e_O$, for a given EH realization. We present the average and the standard deviation of e_G over N_{sim} different simulation realizations in Fig. 4

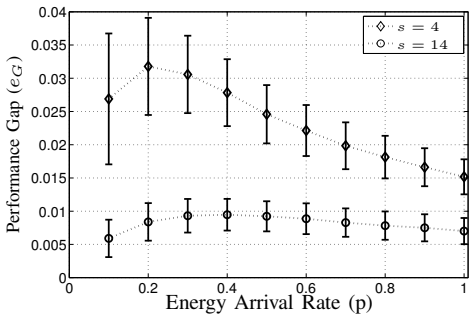


Fig. 4: Performance gap between A_U-n and A_O versus energy arrival rate.

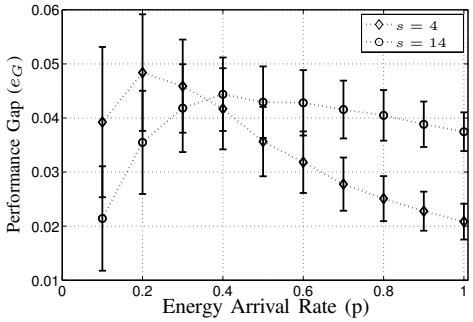


Fig. 5: Performance gap between A_A and A_O versus energy arrival rate.

and Fig. 5 for $X = A_U$-n and $X = A_A$ respectively. Here the deviation is presented with an error bar with a length of one standard deviation on the mean values. Here the scenario with $U = U_a$, $\sigma_h^2 = 1$ is considered whose error values were presented in Fig. 2a and Fig. 3a. We note that, consistent with the presentation of error values, we report the gaps on the normalized error values, i.e. ε/P_x. We observe that for A_U-n, both the mean and the standard deviation are small, illustrating that for most of the EH realizations low-complexity policy A_U-n provides performance relatively close to the optimal. This performance gap is relatively larger for the online strategy A_A which is consistent with the fact that this strategy does not use future channel state information. We note that the performance gap of A_L-n behaves similar to the gap of A_U-n, hence it is not presented here to avoid repetition in the figures.

The power allocations provided by A_O and A_U-n may be different even in the scenarios where the performance is very close. We now provide a scenario that illustrates this. Let x be a low-pass c.w.s.s. with $s = 4$ with $\Lambda_{x,s} = \frac{P_x}{\text{tr}[\Lambda]}\Lambda$, $\Lambda = \text{diag}(\alpha_k)$, $\alpha_k = 1, 0 \leq k \leq s-1$ and $|h_t|^2 = 10$, $\sigma_w^2 = 0.1$. Let $E_t = 1$, for $t = 4k$, $1 \leq k \leq s$ and zero otherwise. By Corollary 3.3, the uniform allocation over equidistant samples, i.e. $a_t = 1$ for $t = 4k$, $1 \leq k \leq s$ and zero otherwise is an optimal strategy. On the other hand, A_U-n provides the most majorized strategy which is given by $a_t = 0.25$ for $4 \leq t \leq n-1$, $a_n = 1$ and zero otherwise. These allocations result in a normalized error of approximately 9.9×10^{-3} and 1.14×10^{-2} for A_O and A_U-n, respectively.

We now discuss the numerical efficiency of the sub-optimal approaches of Section IV. The average computational time of

TABLE I:
Normalized Average Computational Time

	A_O	A_U-2	A_U-$n/2$	A_U-n
$n = 16$	1	2.64	0.70	0.41
$n = 32$	5.42	5.19	0.82	0.55
$n = 64$	80.50	10.32	1.11	0.85

A_U together with that of A_O is provided in Table I for $n = s$, $p = 0.3$. The optimization problem solved by A_L has the same structure as the one for A_U, hence it leads to similar values and is omitted. In Table I, the values are normalized with the value for A_O with $n = 16$. We observe that although the computational time increases for all approaches with increasing n, this effect is most prominent for the approach that directly solves the optimization problem in (6) i.e. A_O. Comparing the computational time for A_U-l_w for different values of l_w, the total time is observed to be higher with small l_w compared to $l_w = n$. This is due to usage of smaller length windows which requires n/l_w calls to the optimization procedure. It is observed that A_U becomes the most numerically efficient approach for all l_w with increasing n. We observe that as n increases, the gap between the computational time values for the direct optimization approach A_O and the sub-optimal approach of A_U-n increases significantly. Together with the close performance of A_U-n to A_O, this supports the usage of A_U-n as a possibly sub-optimal but nevertheless a numerically efficient approach.

VII. Conclusions

We have focused on the remote estimation of a time-correlated signal using an EH sensor. We have considered the problem of optimal power allocation at the sensor under energy causality constraints in order to minimize the MSE at the fusion center. Contrary to the traditional line of work, the correlation between the signal values was an important aspect of our formulation. We have provided structural results for the optimal power allocation strategies for a number of scenarios. In the case of circularly wide sense stationary signals, we have showed that the optimal strategy can be characterized as a water-filling solution for sampled low-pass signals for a fading channel. We have showed that the most majorized power allocation strategy, i.e. the strategy where the power allocation is as balanced as possible, is optimal regardless of the degree of correlation in the case of c.w.s.s. signals with a static correlation coefficient and in the case of sampled low-pass c.w.s.s. signals for a static channel. These results provided important insights into remote estimation of correlated signals under EH constraints that cannot be obtained by considering uncorrelated signals.

We have proposed low-complexity policies for the general case based on upper and lower bounds on the mean-square error. Numerical evaluations have illustrated the performance of low-complexity and optimal policies. The promising performance of the low-complexity approaches and the improvements offered by these approaches in terms of computational time, support the usage of these low-complexity policies as

promising, possibly sub-optimal but nevertheless numerically efficient strategies.

VIII. Appendix

A. Proof of Lemma 3.4

We note that in the parameter estimation case, minimizing $\varepsilon(A)$ is equivalent to maximizing the sum $\sum_{t=1}^{n}|h_t|^2\sigma_{x_t}^2 a_t$. We first consider the case without the energy causality constraints, i.e.

$$\max_{J_t}\sum_{t=1}^{n}|h_t|^2 J_t \tag{40}$$

subject to $\sum_{t=1}^{n} J_t = E_{tot}$, $J_t \geq 0$ where $J_t = \sigma_{x_t}^2 a_t$. The optimal strategy for (40) is given as follows: $q_{t^*} = E_{tot}$, $t^* = \arg\max_{1\leq t\leq n}|h_t|^2$, and $J_t = 0$, if $t \neq t^*$. Hence the optimal strategy is in the form of transmission with all the available energy in the slot with the highest gain. Optimality of this strategy can be seen, for instance, by observing that any other strategy will achieve a smaller objective function since $|h_t|^2 \leq |h_{t^*}|^2$ for $t \neq t^*$. We note that if different time slots have the same maximum channel gain, i.e $|h_{t^*}|^2 = |h_{t_1}|^2 = |h_{t_2}|^2$, $t_1 \neq t_2$, the energy can be allocated arbitrarily between these time slots.

We now go back to the original setting of Lemma 3.4 with the energy causality constraints. We observe that at the first iteration, the procedure gives the optimal possible allocation for the energy allocation up to time t^*. We also observe that one cannot improve the objective function by saving some of this energy for future transmissions since $|h_t|^2 \leq |h_{t^*}|^2$ for $t > t^*$. Similar to the previous case, if we have $|h_{t^*}|^2 = |h_{t_1}|^2 = |h_{t_2}|^2$, $t_1 \neq t_2$, the energy saved up to $t = \max(t_1, t_2)$ can be allocated to the transmissions at t_1 and t_2 in an arbitrary manner (under the condition energy causality constraints are not violated) without any change in the objective function. Thus, at any iteration i, Step-iii of Lemma 3.4 provides an optimal allocation up to t^* at that iteration. Hence the procedure given in Lemma 3.4 provides an optimal strategy.

B. Proof of Lemma 3.5

Let $R_A = \frac{P_x}{s} U_\Omega^\dagger \operatorname{diag}(a_t) U_\Omega$. We observe that

$$\begin{aligned}
\operatorname{tr}[R_A] &= \frac{P_x}{s}\operatorname{tr}[U_\Omega^\dagger \operatorname{diag}(a_t) U_\Omega], && (41)\\
&= \frac{P_x}{s}\operatorname{tr}[\operatorname{diag}(a_t) U_\Omega U_\Omega^\dagger], && (42)\\
&= \operatorname{tr}[\operatorname{diag}(a_t) K_{\mathbf{x}}], && (43)\\
&= \sum_{t=1}^{n} a_t \sigma_{x_t}^2, && (44)\\
&= E_{tot}, && (45)
\end{aligned}$$

where we have used $\operatorname{tr}[AB] = \operatorname{tr}[BA]$ for matrices with appropriate dimensions in (42), $K_{\mathbf{x}} = \frac{P_x}{s} U_\Omega U_\Omega^\dagger$ in (43) and (22b) in (45). We now consider the error expression

$$\varepsilon(A) = \operatorname{tr}\left[\left(\frac{s}{P_x} I_s + \gamma U_\Omega^\dagger \operatorname{diag}(a_t) U_\Omega\right)^{-1}\right], \tag{46}$$

$$\begin{aligned}
&= \sum_{i=1}^{s} \frac{1}{1+\gamma\lambda_i(R_A)} \frac{P_x}{s}, && (47)\\
&\geq \sum_{i=1}^{s} \frac{1}{1+\gamma\frac{\operatorname{tr}[R_A]}{s}} \frac{P_x}{s}, && (48)
\end{aligned}$$

where $\lambda_i(R_A)$ denotes the eigenvalues of R_A. Since (47) is a Schur-convex function of $\lambda_i(R_A)$, it is lower bounded by (48) which is the error associated with a uniform eigenvalue distribution for R_A, i.e. $\lambda_i(R_A) = \operatorname{tr}[R_A]/s = E_{tot}/s$, $i = 1, \ldots, s$. This lower bound in (48) is achievable by choosing $a_t = E_{tot}/P_x$. In particular, this choice of a_t results in $\lambda_i(R_A) = E_{tot}/s$, since $R_A = (P_x/s) U_\Omega^\dagger \operatorname{diag}(E_{tot}/P_x) U_\Omega = (E_{tot}/s) I_s$ where we have used $U_\Omega^\dagger U_\Omega = I_s$.

C. Proof of Lemma 3.6

We first recall that a function of n variables whose value does not change for any permutation of the input is called (permutation) symmetric [39]. We rewrite $\varepsilon(A)$ to show it is a symmetric function of $a_1, \ldots, a_n$ as follows

$$\begin{aligned}
\varepsilon(A) &= \operatorname{tr}\left[(\bar\beta I_n + \bar\alpha e_j e_j^\dagger + \gamma F^{n\dagger} \operatorname{diag}(a_t) F^n)^{-1}\right], && (49)\\
&= \operatorname{tr}\left[R^{-1} - \frac{R^{-1}\bar\alpha e_j e_j^\dagger R^{-1}}{1+\bar\alpha e_j^\dagger R^{-1} e_j}\right], && (50)\\
&= \operatorname{tr}\left[R^{-1}\right] - \frac{\bar\alpha e_j^\dagger R^{-2} e_j}{1+\bar\alpha e_j^\dagger R^{-1} e_j}, && (51)
\end{aligned}$$

where $\bar\alpha = 1/(\alpha+\beta) - 1/\beta$, $\bar\beta = 1/\beta > 0$ and

$$R = \bar\beta I_n + \gamma F^{n\dagger} \operatorname{diag}(a_t) F^n = F^{n\dagger} \operatorname{diag}(\bar\beta + \gamma a_t) F^n.$$

Here (50) follows from the Sherman-Morrison-Woodbury identity with $1 + \bar\alpha e_j^\dagger R^{-1} e_j \neq 0$ [34] and (51) follows from $\operatorname{tr}[AB] = \operatorname{tr}[BA]$ for matrices with appropriate dimensions. Let $\theta_t = 1/(\bar\beta + \gamma a_t)$, hence $R = F^{n\dagger} \operatorname{diag}(1/\theta_t) F^n$ and $R^{-1} = F^{n\dagger} \operatorname{diag}(\theta_t) F^n$. We have

$$[R^{-1}]_{ii} = e_i^\dagger R^{-1} e_i = \sum_{t=1}^{n} \theta_t |[F^n]_{it}|^2 = \frac{1}{n}\sum_{t=1}^{n}\theta_t \tag{52}$$

and similarly $[R^{-2}]_{ii} = (1/n)\sum_{t=1}^{n}\theta_t^2$. Hence we obtain

$$\varepsilon(A) = \sum_{t=1}^{n}\theta_t - \frac{\bar\alpha}{1+\bar\alpha\frac{1}{n}\sum_{t=1}^{n}\theta_t} \frac{1}{n}\sum_{t=1}^{n}\theta_t^2. \tag{53}$$

Here (53) reveals that $\varepsilon(A)$ is a symmetric function of $a_1, \ldots, a_n$. Since $\varepsilon(A)$ is also a convex function of a_t, (due to, for instance, (49) and the fact that $\operatorname{tr}[X^{-1}]$ is convex for $X \succ 0$) $\varepsilon(A)$ is Schur-convex by [39, Ch.3-Prop.C2]. The result follows from Lemma 3.3.

D. Proof of Lemma 4.1

Let $d_t = |h_t|^2 a_t \ \forall t$. We have

$$\begin{aligned}
\varepsilon(A) &= \frac{P_x}{s}\left(\operatorname{tr}\left[(I_s + \gamma\frac{P_x}{s} U_\Omega^\dagger \operatorname{diag}(d_t) U_\Omega)^{-1}\right]\right), && (54)\\
&= \frac{P_x}{s}\left(\operatorname{tr}\left[(I_n + \gamma \operatorname{diag}(d_t) U_\Omega \frac{P_x}{s} U_\Omega^\dagger)^{-1}\right] + s - n\right), && (55)
\end{aligned}$$

$$= \frac{P_x}{s}\left(\mathrm{tr}\left[(I_n + \gamma\,\mathrm{diag}(d_t)K_{\mathbf{x}})^{-1}\right] + s - n\right), \quad (56)$$

$$\geq \frac{P_x}{s}\left(\sum_{t=1}^{n}\frac{1}{1+\gamma|h_t|^2 a_t \sigma_{x_t}^2} + s - n\right). \quad (57)$$

The equality in (55) follows from the equivalence of the non-zero eigenvalues of the matrix products AB and BA; see, for instance, [39, Ch9-A.1.a]. The inequality in (57) is due to the fact that for a p.s.d. matrix $R \in \mathbb{C}^{s\times s}$, $\mathrm{tr}[R^{-1}] \geq \mathrm{tr}[\mathrm{diag}([R]_{ii})^{-1}]$, which, for instance, follows from applying Cauchy-Schwarz inequality $|u^\dagger v|^2 \leq \|u\|^2\|v\|^2$ to $e_i^\dagger R^{-1/2}R^{1/2}e_i = 1$ with $u = R^{-1/2}e_i$, $v = R^{1/2}e_i$ where $e_i \in \mathbb{R}^n$ denotes the ith unit vector.

REFERENCES

[1] D. Gündüz, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designing intelligent energy harvesting communication systems," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 210–216, 2014.

[2] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy Harvesting Wireless Communications: A Review of Recent Advances," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 360–381, Mar. 2015.

[3] O. Ozel and S. Ulukus, "Achieving AWGN capacity under stochastic energy harvesting," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6471–6483, Oct 2012.

[4] Y. Dong, F. Farnia, and A. Özgür, "Near optimal energy control and approximate capacity of energy harvesting communication," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 540–557, March 2015.

[5] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1732–1743, Sept. 2011.

[6] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 1180–1189, March 2012.

[7] J. Yang, O. Ozel, and S. Ulukus, "Broadcasting with an energy harvesting rechargeable transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 571–583, February 2012.

[8] M. A. Antepli, E. Uysal-Biyikoglu, and H. Erkal, "Optimal Packet Scheduling on an Energy Harvesting Broadcast Link," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1721–1731, Sept. 2011.

[9] J. Yang and S. Ulukus, "Optimal packet scheduling in a multiple access channel with energy harvesting transmitters," *Journal of Communications and Networks,*, vol. 14, no. 2, pp. 140–150, 2012.

[10] I. Bahceci and A. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. Commun.*, vol. 56, pp. 1146–1156, July 2008.

[11] A. Shirazinia, S. Dey, D. Ciuonzo, and P. S. Rossi, "Massive MIMO for decentralized estimation of a correlated source," *IEEE Trans. on Signal Process.*, vol. 64, pp. 2499–2512, May 2016.

[12] A. Özçelikkale, H. M. Ozaktas, and E. Arıkan, "Signal recovery with cost constrained measurements," *IEEE Trans. Signal Process.*, vol. 58, pp. 3607–3617, July 2010.

[13] A. Özçelikkale, S. Yüksel, and H. Ozaktas, "Unitary precoding and basis dependency of MMSE performance for Gaussian erasure channels," *IEEE Trans. Inf. Theory*, vol. 60, pp. 7186–7203, Nov 2014.

[14] O. Orhan, D. Gündüz, and E. Erkip, "Source-channel coding under energy, delay, and buffer constraints," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 3836–3849, July 2015.

[15] Y. Zhao, B. Chen, and R. Zhang, "Optimal power management for remote estimation with an energy harvesting sensor," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 6471–6480, Nov. 2015.

[16] M. Nourian, S. Dey, and A. Ahlen, "Distortion Minimization in Multi-Sensor Estimation With Energy Harvesting," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 524–539, Mar. 2015.

[17] S. Knorn, S. Dey, A. Ahlen, and D. E. Quevedo, "Distortion Minimization in Multi-Sensor Estimation Using Energy Harvesting and Energy Sharing," *IEEE Trans. Signal Process.*, vol. 63, pp. 2848–2863, June 2015.

[18] A. Nayyar, T. Başar, D. Teneketzis, and V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Trans. Autom. Control*, vol. 58, pp. 2246–2260, Sept 2013.

[19] M. Nourian, A. Leong, and S. Dey, "Optimal energy allocation for Kalman filtering over packet dropping links with imperfect acknowledgments and energy harvesting constraints," *IEEE Trans. Autom. Control*, vol. 59, pp. 2128–2143, Aug 2014.

[20] M. Calvo-Fullana, J. Matamoros, and C. Anton-Haro, "Reconstruction of Correlated Sources with Energy Harvesting Constraints," in *European Wireless Conf. 2015*, pp. 1–6, 2015.

[21] R. Gangula, D. Gündüz, and D. Gesbert, "Distributed compression and transmission with energy harvesting sensors," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1139–1143, 2015.

[22] C. Tapparello, O. Simeone, and M. Rossi, "Dynamic Compression-Transmission for Energy-Harvesting Multihop Networks With Correlated Sources," *IEEE/ACM Transactions on Networking*, vol. 22, pp. 1729–1741, Dec. 2014.

[23] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1147–1158, May 2003.

[24] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, 2008.

[25] A. Özçelikkale, T. McKelvey, and M. Viberg, "Transmission strategies for remote estimation under energy harvesting constraints," in *Proc. European Signal Process. Conf. (EUSIPCO)*, pp. 572 – 576, 2016.

[26] S. Guo, C. Wang, and Y. Yang, "Joint Mobile Data Gathering and Energy Provisioning in Wireless Rechargeable Sensor Networks," *IEEE Trans. on Mobile Computing*, vol. 13, pp. 2836–2852, Dec. 2014.

[27] R. Du, C. Fischione, and M. Xiao, "Lifetime maximization for sensor networks with wireless energy transfer," in *Proc. IEEE Inter. Conf. on Communications (ICC)*, pp. 1–6, 2016.

[28] Q. Wang and M. Liu, "When simplicity meets optimality: Efficient transmission power control with stochastic energy harvesting," in *Proc. of IEEE INFOCOM*, pp. 580–584, April 2013.

[29] A. Özçelikkale, T. McKelvey, and M. Viberg, "Performance bounds for remote estimation with an energy harvesting sensor," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pp. 460–464, 2016.

[30] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.

[31] R. M. Gray, *Toeplitz and Circulant Matrices: a Review*. Now Publishers Inc., 2006.

[32] K.-H. Lee and D. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. 24, pp. 1283–1290, Dec 1976.

[33] B. D. O. Anderson and J. B. Moore, *Optimal filtering*. Prentice-Hall, 1979.

[34] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Review*, vol. 23, no. 1, pp. 53–60, 1981.

[35] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming*, vol. 95, no. 2, pp. 189–217, 2003.

[36] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[37] CVX Research Inc., "CVX: Matlab software for disciplined convex programming 2.0." http://cvxr.com/cvx, 2012.

[38] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[39] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. Academic Press, 1979.

[40] A. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, 2013.

[41] Z.-Q. Luo, W.-K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, pp. 20–34, May 2010.

[42] F. Mezzadri, "How to generate random matrices from the classical compact groups," *Notices of the AMS*, vol. 54, pp. 592 – 604, 2007.